1. Which statements are usually true of a left-skewed distribution? (circle all that are correct)

(a) The mean is smaller than the median.
(c) $Q_3$ is closer to the median than $Q_1$.

2. Which of the following statements about the slope $b_1$ are true (circle all that are correct):

(a) A value of $b_1$ exactly equal to 0 indicates no linear association.
(b) The slope depends on the measurement units of both the explanatory and response variables.
(c) If $b_1 < 0$, then the direction of association between explanatory and response variables is negative.


$$z = \frac{100 - 90}{25} = .4.$$  

The normal table shows that the area to the left of $-0.4$ is 0.3446.

4. What is the 30th percentile of the N(90,25) distribution?

The 30th percentile of the $(N(0,1))$ distribution is $-0.52$ (from the normal table). The 30th percentile for the $N(90,25)$ distribution is obtained as follows

$$-0.52 = z = \frac{x - 90}{25} \Rightarrow x = -0.52 * 25 + 90 = 77.$$  

5. What percent of males were admitted?

$$\frac{120}{120 + 205} = .369 \Rightarrow 36.9\%.$$  

6. What percent of those rejected were male?

$$\frac{205}{205 + 391} = .3439 \Rightarrow 34.4\%.$$  

7. Find the marginal distribution of admittance outcome.

$$\frac{120 + 202}{918} = .351 \quad \frac{205 + 391}{918} = .649.$$  

1
(4 pts) 8. Find the conditional distribution of gender outcome for those who were rejected.

<table>
<thead>
<tr>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>205</td>
<td>596</td>
</tr>
<tr>
<td>344</td>
<td>656</td>
</tr>
</tbody>
</table>

The following data set consists of a random sample of $n = 11$ observations from the integers $1, 2, \ldots, 100$. The observations have been sorted from smallest to largest.

<table>
<thead>
<tr>
<th>$i$</th>
<th>$x_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>43</td>
</tr>
<tr>
<td>5</td>
<td>49</td>
</tr>
<tr>
<td>6</td>
<td>63</td>
</tr>
<tr>
<td>7</td>
<td>67</td>
</tr>
<tr>
<td>8</td>
<td>81</td>
</tr>
<tr>
<td>9</td>
<td>85</td>
</tr>
<tr>
<td>10</td>
<td>92</td>
</tr>
<tr>
<td>11</td>
<td>95</td>
</tr>
</tbody>
</table>

(5 pts) 9. Find the median. $M = 63$.

(5 pts) 10. Find the first and third quartiles. $Q_1 = \frac{19+43}{2} = 31$. $Q_3 = \frac{81+84}{2} = 82.5$.

(5 pts) 11. Find the IQR. $IQR = 82.5 - 31 = 51.5$.

(9 pts) 12. The histogram to the right shows the distribution of values in a data set.

(a) Provide the names of measures of center and spread appropriate for these data.

Median and IQR.

(b) Provide the names of measures of center and spread not appropriate for these data.

Mean and standard deviation.

The National Health and Nutrition Examination Survey reports that for female US residents 21 to 30 years old, the mean body weight is $\mu = 149$ lb and the standard deviation is $\sigma = 37$ lb.

(6 pts) 13. Assuming that the distribution is normal, what proportion of weights are greater than 186 lb?

$$z = \frac{186 - 149}{37} = 1 \Rightarrow .1597$$

From the normal table, or .16 using the 68-95-99.7 rule.

(6 pts) 14. Using the 68-95-99.7 rule, find the 97.5th percentile of the weight distribution. The 97.5th percentile is found 2 standard deviations above the mean. Thus, the answer is

$$\mu + 2\sigma = 149 + 2 \times 37 = 223.$$

The National Health and Nutrition Examination Survey also reported the following percentiles of the weight distribution.

<table>
<thead>
<tr>
<th>Percentile</th>
<th>5th</th>
<th>10th</th>
<th>15th</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>85th</th>
<th>90th</th>
<th>95th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight (lb)</td>
<td>106</td>
<td>112</td>
<td>119</td>
<td>128</td>
<td>150</td>
<td>178</td>
<td>198</td>
<td>218</td>
<td>240</td>
</tr>
</tbody>
</table>
15. To the extent possible, draw a boxplot using these summary statistics. It’s not possible to draw all features of the boxplot using these summary statistics. Instead of drawing the whiskers, draw the fences. It’s not possible to identify outliers using these summary statistics, so ignore that feature. For those features that can be drawn, **place the numerical value of the statistic that defines the feature.**

First, \( M = 150, Q_1 = 128 \) and \( Q_3 = 178 \) \( \Rightarrow \) IQR \( = 50 \) and \( 1.5 \text{IQR} = 75. \)

\[
\text{L. fence} = 53 \quad \text{U. fence} = 253
\]

16. The following plot shows the relationship between stopping distance and speed of cars (data are from the 1920’s). Which of the following statements are true?

(a) □ The association is moderate and positive.

(b) □ It’s reasonable to think of stopping distance responding to speed.

![Plot showing relationship between speed and stopping distance](image)

17. The following plot shows the relationship between year and world record times for the mile run, 1861-2003, by gender. Draw your best guesses of linear models that summarize the relationships (that is, separate lines for each gender). Which of the following statements are true? Place a check mark in front of each true statement.

(a) □ The association between year and record times is strong, negative, and linear when looking at males alone.

(b) □ There is no sensible interpretation of either intercepts.

![Plot showing relationship between year and gender for world record times](image)

The least squares regression results for the female athletes are given below.
Table 1: World record times for the mile run, 1861-2003. Summary statistics: regression of time on year.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_0$</td>
<td>2309.42</td>
<td>221.0</td>
<td>10.448</td>
<td>0.00000005</td>
</tr>
<tr>
<td>$b_1$</td>
<td>-1.0337</td>
<td>0.1117</td>
<td>-9.256</td>
<td>0.00000025</td>
</tr>
</tbody>
</table>

Residual standard error: 3.269 on 14 degrees of freedom Multiple R-squared: 0.8595, Adjusted R-squared: .8495 F-statistic: 85.67 on 1 and 14 DF, p-value: 0.00000025

(8 pts) 18. Write the equation of the least squares lines in the context of the problem.

\[
\text{record time} = 2309.4 - 1.034\text{year}
\]

(2 pts) 19. Interpret the slope in the context of the problem.

(a) ☐ For every 1 additional year, record times decrease by 1.0337 seconds.

(4 pts) 20. Identify which statements are correct interpretations of $R^2 = .8595$.

(a) ☐ 85.95% of the variation in record times is explained by year.